Selecting Attributes to Measure the Achievement of Objectives

by

Ralph L. Keeney[a] and Robin S. Gregory[b]

Abstract

The foundation for any decision is a clear statement of objectives. Attributes clarify the meaning of each objective and are required to measure the consequences of different alternatives. Unfortunately, insufficient thought typically is given to the choice of attributes. This paper addresses this problem by presenting theory and guidelines for identifying appropriate attributes. We define five desirable properties of attributes: they should be unambiguous, comprehensive, direct, operational, and understandable. Each of these properties is discussed and illustrated with examples, including several cases in which one or more of the desirable properties are not met. We also present a decision model for selecting among the different types of natural, proxy, and constructed attributes.

August, 2004

In press, *Operations Research*

[a] Fuqua School of Business, Duke University, Durham, NC 27708 (keeney@duke.edu)

[b] Decision Research, Eugene, OR 97401 (rgregory@interchange.ubc.ca)

1. Introduction

We routinely make decisions in organizations, businesses, and our personal lives. Most of these are made intuitively without analysis to facilitate clear thinking. Sometimes, an analysis is useful to provide insight about the relative desirability of alternatives in order to enhance the quality of the choice.

The foundation for any analysis is the set of objectives considered and the set of alternatives for achieving those objectives. To describe the consequences of alternatives and make value tradeoffs between achieving relatively more or less on different objectives, it is necessary to identify a measure for each objective. We refer to such a measure as an attribute. The terms performance measure, criterion, and metric are often used as synonyms. This paper presents theory, guidelines, and examples for identifying appropriate attributes to measure the achievement of objectives in analyses.

A simple example illustrates our concern. Suppose a state fisheries agency wants to improve the salmon run on a particular river and several alternatives have been suggested. The stated objectives, common to initiatives of this type, are to maximize the increase in salmon spawning areas and to minimize the economic cost of the project. In order to compare projects, it is necessary to understand how much salmon spawning areas are improved by the various alternatives and the cost of each alternative. Measurement of these two objectives is therefore important. The attribute "dollars of cost" seems obvious for the objective minimize economic cost. However, the choice of an attribute for measuring increase in salmon spawning areas is not obvious. Maybe it should be the river length of new spawning areas created, or the number of distinct locations where spawning areas are created, or some attribute meant to capture the expected productivity of new and existing spawning areas, or even an attribute or attributes that collectively deal with increased spawning areas and increases in their productivity.

Suppose that the attribute "additional feet of new spawning areas created in the river" is chosen. Then assume that, after appraising alternatives, two are judged to be better than the others. One alternative costs $3 million and creates 2000 feet of additional spawning habitat, whereas the second costs $5 million and creates 4000 of additional spawning habitat. The first alternative has a cost of $1.5 thousand for each additional foot and the second alternative costs $1.25 thousand for each additional foot. Therefore, without careful thought, the latter alternative may appear to be the better choice. However, cost effectiveness, measured in cost per foot of spawning area created, is not necessarily an appropriate measure. What if the additional 2000 feet of spawning area created is more than sufficient for all the salmon that could survive in this river after spawning? In this case, 2000 feet and 4000 feet of spawning area would essentially be equivalent, so it would be a poor decision to spend an additional $2 million when it provides essentially nothing of value. It is important that the attributes be identified such that one can responsibly make the value tradeoffs necessary at this balancing stage of evaluation.

Use of the attribute "additional feet of new spawning areas created in the river" raises other fundamental issues. This measure treats all areas of the river as identical, but is this an appropriate assumption? Perhaps new spawning areas are not essential, instead existing areas (not covered by the current measure) should be restored. Perhaps areas outside of the river -- in tributaries or side channels– could usefully be improved; again these are not covered under the existing measure. A further source of confusion is that some terms may not be defined clearly: not all participants may agree on the definition of a "new" spawning area, for example, or what is included as being "in the river".

2. Fundamental Concepts about Attributes

Previous research has identified three different types of attributes: natural attributes, constructed attributes, and proxy attributes (Keeney, 1992). In some cases, an attribute may look more like a hybrid of two of these types, but this trichotomy is useful for discussing features of attributes.

Natural attributes are in general use and have a common interpretation. In the simple example of section 1, the objective "minimize cost" has the natural attribute "cost measured in dollars". In another problem, if one objective is to minimize the loss of wildlife habitat, a natural attribute might be "acres of lost habitat". For decisions concerning life and death, such as setting automobile speed limits, the objective of minimizing fatalities might be measured by the natural attribute "number of fatalities". Basically, most natural attributes can be counted or physically measured.

They also have the property that they directly measure the degree to which an objective is met.

Proxy attributes share many qualities of natural attributes. They usually involve a scale that is in general use that can be counted or physically measured. The difference is that they do not directly measure the objective of concern. For a decision involving setting speed limits, for example, one proxy attribute for the objective "minimize fatalities" is the "number of vehicle accidents". Certainly the number of vehicle accidents is related to the number of fatalities, but it does not directly measure those fatalities. A proxy attribute is less informative than a natural attribute because it indirectly indicates the achievement of an objective.

Proxy attributes typically are used when it is difficult to select a natural attribute to measure an objective. Consider a manufacturing firm with the objective to "maximize quality" for a given product. There may be no obvious natural attribute to measure this objective. However, an available proxy attribute is the percentage of sales that are returned. This attribute clearly does not directly measure quality. It also neglects customers who are disappointed with the product quality but do not return it. Yet, they may select a competitor's product next time. For reasons such as this, it is difficult to interpret the significance of different levels of a proxy attribute.

Constructed attributes are sometimes developed to measure directly the achievement of an objective when no natural attributes exist. For instance, suppose members of the public living in the vicinity of a proposed toxic waste site are fearful of the health implications to their families and friends. One objective would be to reduce the amount of fear felt by members of the public. Simply using the number of people experiencing fear may be inadequate, because the fear levels of different people may vary significantly. Yet no natural scale exists to measure fear levels. As a result, a scale may need to be constructed out of several aspects typically associated with fear, perhaps including physical measures (e.g., blood pressure) or behaviors (e.g., people who stop drinking local tap water in the area). Another objective in that same decision may be to increase local public support for siting the facility. A possible natural attribute would be the percentage of the public supporting the facility or opposed to the facility. A constructed scale for this situation might be as illustrated in Table 1, which focuses on attitudes and actions of various groups of the public. Such a constructed scale identifies two or more distinct levels and defines them with an appropriate description.

Once a constructed attribute has been commonly used in practice, many people become familiar with it and it begins to take on properties of a natural attribute. The Dow Jones Industrial Average was introduced in 1896 and expanded in 1928 to include the prices of 30 stocks to measure the movement of the stock market. This originally constructed attribute is now more like a natural attribute to individuals knowledgeable about the stock market. The same process is occurring today with more broadly based constructed attributes of market behavior such as the S&P 500 and the Russell 1000. It is familiarity with an attribute and the ease of interpreting the attribute levels that distinguishes natural attributes from constructed attributes.

It is useful to recognize that the concept of an attribute has two associated notions, one qualitative and one quantitative. These can best be illustrated by example. For an objective such as minimize cost, we might define the attribute to be cost in dollars. The qualitative notion is "cost" and “dollars” is a quantitative scale for cost. In this case, they almost seem like the same thing and naturally go together.

However, even in this "apparently obvious" case, important issues for applications arise. In a key study done with and for the Department of Energy (DOE) to evaluate the final candidate sites for the U.S. high-level nuclear waste repository (Merkhofer and Keeney, 1987), one major objective was "minimize repository costs." Base-case estimates for the five sites ranged from $7.5 billion to $12.9 billion in 1987 dollars. In doing the analysis, it seemed as though policy makers in DOE did not consider the cost differences to be very significant. Numbers like 7.5 and 12.9 were conceptually small and the difference seemed even smaller. To make these numbers salient within DOE and to legislators, politicians, and the public who would later see the report, the analysts chose to use the scale "millions of dollars" rather than billions in the study. It is perhaps a little awkward to always say and write numbers like 7,500 millions of dollars and 12,900 millions of dollars, but they did communicate better the magnitude of the cost estimate. They look and sound big, as they are.

In another decision, one objective might be to maximize fuel efficiency of automobiles. Here, the qualitative notion is "mileage", which can be measured by different quantitative scales. The scale typically used by an individual in the US is miles per gallon, so we probably would define the attribute as mileage measured in miles per gallon. However, for indicating the fleet fuel efficiency of automobile manufacturers, the preferred measure might be gallons per mile. In Europe, the typical natural attribute for the same objective is mileage measured in liters per hundred kilometers. However, note that more miles per gallon is better whereas less liters per hundred kilometers is better.

3. Desirable Properties of Attributes

The foundation underlying any analysis is the set of objectives. There are desired properties of this collective set of objectives that, when possessed, can greatly enhance the value of any subsequent analysis (Keeney, 1992). However, the objectives underlying an analysis are too often neither well thought out nor appropriate. In such cases, even great attributes for those objectives will not make up for this inadequacy, and any analysis will provide much less insight than would be the case were a set of logical and complete objectives in place. When the objectives do provide a good foundation for describing consequences, then inadequate attributes can seriously damage an analysis. Common errors include attributes that are ambiguous (and therefore are interpreted differently by different individuals), fail to take advantage of the available information relating to consequences, or incompletely describe the consequences of the objective they are intended to measure. Our experience is that, even with important and highly visible decision processes, insufficient thought typically is given to the identification and choice of attributes. This paper is not specifically concerned with the properties of the set of objectives. It focuses on the desired properties of attributes selected for already specified objectives.

In previous works (Keeney, 1982, 1992; Gregory and Failing, 2002), we have discussed some desirable properties to take into account in developing good attributes. Here, we extend this work and specify five desirable properties of good attributes. The previous work concerned necessary properties for good attributes, whereas this work presents a sufficient set of properties for good attributes. These five properties, with simple definitions of each, are as follows:

Unambiguous—A clear relationship exists between consequences and descriptions of consequences using the attribute,
Comprehensive--The attribute levels cover the range of possible consequences for the corresponding objective and value judgments implicit in the attribute are reasonable,
Direct--The attribute levels directly describe to the consequences of interest,
Operational—In practice, information to describe consequences can be obtained and value tradeoffs can reasonably be made,
Understandable—Consequences and value tradeoffs made using the attribute can readily be understood and clearly communicated.

There are several interrelationships among these five properties. If an attribute is ambiguous, it almost for sure will fall short in terms of being comprehensive or understandable. If an attribute is not comprehensive or not direct, it will be much less operational than otherwise. If an attribute is not operational, then of course there is not a good understanding of the consequences. And if an attribute is not understandable, then it is naturally not very operational and likely ambiguous.

<u>Unambiguous</u>. An attribute is unambiguous when there is a clear relationship between the consequences that might or will occur and the level of the attribute used to describe those consequences. Unambiguous attributes must be neither vague nor imprecise. Consider the objective minimize cost and the attribute cost in millions of dollars. If the cost of one alternative is $16.3 million, analysts would simply describe that consequence as $16.3. On the other hand, if the attribute only vaguely categorized costs as high, medium, or low, it might not be obvious which attribute level is appropriate for $16.3 million. Different people could interpret high, medium, and low differently without specific ranges in dollars used to guide interpretation of these terms.

Even if this vagueness were eliminated by defining medium cost as $10-20 million for a specific decision, such an attribute would still be ambiguous. Although $16.3 million is clearly categorized as medium, the description is unnecessarily imprecise. People interpreting a medium cost consequence would only know that the cost is in the range of $10-20 million. They would not know whether the cost was $11 million, $19 million, or $16 million. Yet, there likely is a significant difference in the desirability of consequences of $19 million and $11 million. There is no reason for loss of such useful information in describing and evaluating alternatives.

Another shortcoming arises when uncertainties are involved, which is usually the case with important decisions. Suppose the cost of a particular alternative was described with a probability distribution that ranged from \$18 million to \$23 million. Furthermore, suppose medium was defined as \$10-20 million and high was defined as \$20 million and above. Would the consequences be categorized as medium or high cost? Of course, one might say there is a 40% chance of a medium cost and a 60% chance of a high cost in this situation, but that still does not address the issue of not knowing exactly what the medium costs or the high costs might be.

In summary, an unambiguous attribute has the properties that when you know what the consequence is or will be, you know exactly how to describe it using the attribute, and when you know the description of a consequence in terms of the attribute level, you know exactly what the corresponding consequence is or will be. With uncertainties present, a full description of consequences with respect to an objective is given by a probability distribution over the associated attribute. This issue of uncertainty is completely different from that of ambiguity (Ellsberg, 1961; Camerer and Weber, 1992). Nevertheless, it seems as though ambiguous attributes are sometimes chosen to disguise or to avoid hard thinking about the uncertainties.[1]

<u>Comprehensive</u>. An attribute is comprehensive if its attribute levels cover the full range of possible consequences and if any implicit value judgments that are part of the attribute are appropriate for the decision problem being addressed.

Consider the decision of setting a national ambient air quality standard for carbon monoxide. One of the fundamental objectives is to minimize detrimental health effects from carbon monoxide. Breathing more carbon monoxide increases carboxyhemoglobin and decreases the oxygen carried by the blood, which leads to detrimental health effects. These include fatal and non-fatal heart attacks. Consider the attribute "number of fatalities" for this objective. This attribute does not cover the full range of possible consequences, as all detrimental health effects are not fatal heart attacks.

1

Either a second attribute, such as the number of non-fatal heart attacks, or a composite attribute that includes both non-fatal and fatal consequences is necessary to be comprehensive. In Keeney (1992), four attributes were used to completely describe detrimental health effects. These were number of fatal heart attacks, number of non-fatal heart attacks, number of angina attacks, and number of peripheral vascular attacks.

In some cases, an attribute is not comprehensive because experts want to rely on a more narrow set of measures than is appropriate. Risks that threaten public health and safety, such as exposure to toxic chemicals, provide a well-known example; technical experts frequently want to rely on mortality and morbidity changes whereas public participants want to expand the set of attributes to include concerns such as the voluntariness of exposure or the degree to which scientists are thought to understand the problem (Slovic, 1987). Another example comes from a study conducted for the school board of a large urban area (Gregory, 2003). The goal of the study was to assist the regional school board in allocating funds to local schools on the basis of need in a defensible manner. Experts wanted to rely on average household income as the sole measure of need, whereas public participants sought to expand the set of attributes to include concerns such as the percentage of children living below the poverty line and the mobility of children in schools (i.e., student transfer as a percentage of school enrollment). Including these additional concerns improved the comprehensiveness of the attributes and, in turn, the acceptability of the recommended allocation alternative.

Comprehensiveness also requires that one consider the appropriateness of value judgments embedded in attributes. Whenever an attribute involves counting, such as the number of fatalities, there is the assumption that each of the items counted is equivalent. With the number of fatal heart attacks, there is a built-in assumption that a fatal heart attack for a 45-year-old is equivalent to a fatal heart attack for a 90-year-old. Is this a reasonable value judgment for a particular decision? There isn't a right or wrong answer, but it's an issue that should be considered in selecting the attribute. To be extreme, consider a decision concerning automobile safety. If the objective to minimize fatalities is measured by the attribute number of fatalities, this assumes that the death of a 10-year-old is equivalent to the death of a 90-year-old. Many people think this is inappropriate, as they consider the death of the 10-year-old to be more significant. One way to account for this is to use the attribute "years of life lost" to measure the objective "minimize fatalities". If the expected lifetime of a 10-year-old is 80, then 70 years of life is lost if a 10-year-old dies in an automobile accident. If the expected lifetime of the 90-year-old is 95, then 5 years of life is lost due to a death in an automobile crash. With the attribute of "years of life lost", the death of a 10-year-old would count 14 times as much as the death of a 90-year-old (the ratio of 70 years to 5 years). Although the implications of such value judgments rarely are made explicit, their significance is underscored by the current public policy debate regarding the different values embedded in the two attributes "number of fatalities" and "years of life lost" (Seelye and Tierney, 2003).

Recent research in medical decision-making has taken a further step and developed a scale called "quality adjusted life years," or QALY (Gold et al, 2000). With QALYs, a year of life lost of a healthy individual counts more than a year of life lost of a less healthy individual. Whether the number of fatalities, years of life lost, or quality adjusted years of life lost is a more appropriate attribute in any specific decision problem needs to be appraised using all five desired properties of attributes.

Direct. An attribute is direct when its attribute levels directly describe the consequences for the fundamental objective of interest. A common example where this does not occur is when guidelines have been developed to cover an issue such as worker safety. Attributes measuring how well the guidelines are being met (e.g., what number, or what percentage, of the guidelines have been met, or have been violated, during a specified period) are not direct as they do not directly provide useful information about worker deaths or injuries. More useful information could be provided by attributes that are direct such as the number of worker fatalities, the number of worker injuries of different severity, and the lost-time experienced by workers due to accidents and fatalities. A highly visible example of this shortcoming is the assessment of cleanup at 36 contaminated sites recently undertaken by the U.S. Department of Energy (DOE, 2000). The primary criterion for the assessment was achievement of compliance with regulations governing remediation activities at these sites. However, degree of compliance likely does not have a clear relationship to the consequences of interest, namely public and worker safety or environmental damage.

An attribute that is not direct is sometimes intentionally selected when decision makers seek to distort the results of a decision process. An example of this gaming arises when pronouncements for reductions in the size of government make use of the attribute “number of employees” but fail to provide information about increases in funding going to the new contractors who are hired to replace outgoing employees. Another example occurs when CEOs are rewarded with options on the basis of stock performance, ignoring how the stock price was raised (e.g., through buybacks rather than growth). In situations such as these, an attribute that is not direct may be preferred by decision makers because of their desire to hide controversial implications of their choices or to present potentially troubling or controversial information in the most favorable light.

Operational. Even if attributes are unambiguous, comprehensive, and direct, there is also the practical question of whether they are operational. One aspect of this concerns how easy it is to obtain the information describing consequences. In the widely cited work of the so-called “Montreal Process,” for example, twelve nations with temperate and boreal forests developed a framework that included 67 “indicators” for measuring the success of management policies. This cumbersome framework does not work well for informing decision makers, as quality information is not available for many of this large number of measures. Effort spent trying to collect rough data for a large set of concerns might better be spent learning in more detail about a smaller set of important measures.

Even for attributes that are carefully considered, gathering the information or assessing the judgments to describe consequences may require too much money or time to be operational. In some contexts using the attribute "years of life lost" may create too demanding a task for the analyst. When setting ambient air quality standards, for example, limits in current knowledge and the available resources may render it nearly impossible to make good estimates of how many individuals of different ages might have fatal heart attacks as a function of the level of carbon monoxide in the air. Tradeoffs are always necessary between how easy and practical it is to do an analysis and the additional insight that would be provided if the analysis were done more thoroughly. Sometimes, this boils down to trading off the desired properties for attributes of being comprehensive and operational.

Another aspect of operational attributes concerns whether they enable decision-makers to make informed value tradeoffs concerning how much of one attribute is an even swap for a given amount of a second attribute. These value tradeoffs are necessary to balance the various pros and cons of alternatives. Yet, we recognize that the ability to make informed value tradeoffs requires more than the selection of appropriate attributes. For example, some individuals find making a value tradeoff between economic costs and the potential loss of human life objectionable for moral or ethical reasons. For these individuals, even unambiguous, comprehensive, and direct attributes are not necessarily operational because of their refusal to consider the value tradeoffs explicitly. Tradeoffs such as these, when many people refuse to make them, are referred to as "taboo tradeoffs" (see Fiske & Tetlock, 1996). This unwillingness of people, including some responsible for setting related policies, to make reasoned value tradeoffs poses a dilemma for making informed choices on important decisions.

Understandable. The fifth desirable property is that attributes should be understandable to anyone interested in the analysis. They need to be understandable to those doing the analysis, to decision-makers who want to interpret the analysis, and to stakeholders who want to be informed by the analysis.

Understandability is essential for clear communication. One should be able to easily communicate the pros and cons of various alternatives in terms of their consequences. The standard on understandability for an attribute is that an individual understands the consequences if they are given the attribute levels. With consequences such as cost, number of fatalities, and length of salmon spawning areas created in a river, there would likely be a high level of understandability. Yet for an objective such as maximize mileage of automobiles discussed earlier, the attribute "miles per gallon" has a high level of understandability for people in the United States whereas the attribute "liters per hundred kilometers" is almost incomprehensible. This is the case even though there is a one-to-one relationship between the descriptions of fuel efficiency on these two attributes.

In many cases, technical experts want to rely on attributes that are not understandable to the larger audience of participants concerned about alternatives for dealing with a problem. For instance, with decisions concerning nuclear power or nuclear waste, one objective is often to minimize possible cancer cases to the public due to radioactive emissions from nuclear material. Because of the scientific difficulty of relating radiation exposure to cancer (i.e. the dose-response relationship) and the political sensitivity to talking about radiation-induced cancers, many people avoid using the understandable attribute "number of cancer cases induced". As a result, many studies instead have used the proxy attribute "person-rems of radiation exposure" for the corresponding objective. But what percentage of the public, even the technically trained public, knows whether an exposure of four person-rems is horrendous or insignificant in terms of health implications? Person-rems is inadequate as an attribute in an analysis for public use in terms of understandability.

The understandability of an attribute depends in part on the extent to which it exists within a context that makes sense. Consider a decision to place a dam on some small river to generate power. One objective concerns minimizing the reduction of summer water flows as a result of power generation. An attribute might be the percent of reduced summer flows, and three alternatives may have reductions of 20%, 30%, and 40%. However, the base from which these percents are measured may differ. Furthermore, year-to-year fluctuations in rainfall and snowmelt may naturally cause variations of 100%, 35%, and 60% in the rivers respectively. It would be difficult for most decision-makers and stakeholders to adequately understand the attribute levels because this requires a thorough knowledge of the context for the evaluation. This point is similar to that made by Hsee (1996) in his research on the concept of evaluability. Further, focusing on the reduction in water flows may divert attention from the more fundamental concern about the environmental consequences associated with the predicted post-project water flows.

4. A Decision Model for Selecting Attributes

The selection of an attribute should be viewed as a decision. The quality of a selected attribute, on average, will be better if alternative attributes are considered. This is especially true when there is no obvious attribute, such as dollars to measure an objective concerning cost. The attribute chosen for any objective should be the one that best meets the five desirable properties outlined in section 3. These properties can be thought of as the objectives that one wishes to achieve in decisions involving the choice of attributes.

Each selection of an attribute needs to be addressed in the context of the decision problem being analyzed. That being said, there is still value in looking at the generic problem of what types of attributes are generally best to use in decision problems.

Appraisal of Types of Attributes.

There is strong prescriptive advice for the types of attributes to select. If a natural attribute can be found that is comprehensive, direct, and operational, it should be selected. When that isn't the case, effort should go into construction of an attribute. If there is no time for that effort or if it does not lead to a good constructed attribute, then a proxy attribute should be chosen.

The logic for this ordering is that natural attributes have a general understanding for many people. Information often is collected in terms of natural attributes so it is readily available or can be gathered without inordinate effort. Also, since people commonly think in these terms, it is easier for them to make value tradeoffs. In short, natural attributes are the best way to measure objectives. However, for some important objectives, there are no natural attributes. Examples include objectives that address the social disruption to isolated, small communities when large facilities are built nearby, the morale within a company or organization, the physical pain of different treatments for a disease, or a community's pride in new civic infrastructure.

Construction of attributes should involve individuals knowledgeable about the consequences of concern. This should lead to unambiguous attributes that are comprehensive and direct. If the basis for their construction is adequately described and the manner in which data are gathered and value tradeoffs are made is also clarified, then people should be able to understand and communicate about what was done and why. This typically won't be as easy as with natural attributes, but a well-constructed scale should do a good job of meeting the five desirable properties of attributes.

Only when there is no appropriate natural attribute or constructed attribute should one select a proxy attribute. In this case, it is generally best to choose a proxy attribute with a natural scale, so it likely meets the properties of being unambiguous and comprehensive. Also, it should be selected such that information can be readily gathered to describe consequences, thereby meeting that part of the operational property. A major shortcoming of a proxy attribute is that it does not directly describe the consequences. This, in turn, makes it difficult for decision makers to understand the consequences and assess reasonable value tradeoffs involving the proxy attribute.

Suppose that one is looking at national ambient air quality standards for carbon monoxide, a case discussed earlier. The alternatives are evaluated in terms of the cost to meet different standards and their effects on air quality, measured in parts per million of carbon monoxide. This is a proxy attribute for minimizing health effects. How can one responsibly think about whether it is worthwhile to tighten the standards from four parts to three parts per million of carbon monoxide for an additional cost of $3 billion annually? The answer is that one has to think about the differences in health effects associated with four and three parts per million. However, this complexity is likely what led to the fact that a direct attribute, such as the number of heart attacks, was not used in the first place.

A Flowchart for Selecting Attributes.

Figure 1 outlines the logic for selecting attributes for each of the objectives in an analysis. One first tries to identify natural attributes and selects one if a good choice is found. If not, then one tries to construct an attribute. If neither good natural attributes nor good constructed attributes are available, then a proxy attribute should be chosen.

An important additional aspect in addressing specific decision problems is illustrated in Figure 1. When one tries to identify a natural attribute for a corresponding objective, the difficulty may be that the objective is too broad. At this stage, it is useful to try to decompose that broad objective into component objectives and then identify natural attributes for each of those components. If natural attributes are found for only some of the components, then one proceeds to develop constructed attributes for the other component objectives.

As an example, a major objective of numerous decision problems concerning health is to minimize health impacts to the public. No single adequate natural attribute may be found, because the health effects concern both mortality and morbidity. In this case, the overall objective of minimizing health effects first may be split into minimizing mortality and minimizing morbidity. Then a natural attribute, such as the number of fatalities, may be selected for the mortality objective. However, no adequate natural attribute for the objective of minimizing morbidity may be found. The likely reason is that different types of morbidity range greatly in terms of severity: having a cold and having pneumonia are not equivalent. Again, we may decompose the objective of minimize morbidity into two objectives, such as minimize cases of pneumonia and minimize cases of colds, if those are the two major morbidity conditions. Then each of these might be measured by the natural attribute of number of incidences. Alternatively, it may be decided not to decompose morbidity into specific objectives, especially if many different morbidity effects are possible. In such situations, constructing an attribute for the morbidity consequences, as discussed in Section 6, may be appropriate.

5. Choices among Natural Attributes

For many objectives, there will appear to be an obvious natural attribute. In other cases, there may be different possible natural attributes from which one should be chosen. Each of these natural attributes is likely to satisfy most of the desirable properties of attributes, in that it probably is unambiguous, direct, operational, and understandable. The two aspects of being comprehensive, however, need to be carefully appraised in selecting the attribute. First, does the natural attribute cover the range of possible consequences for the objective? Second, are appropriate value judgments inherent in the attribute? We illustrate these concerns with three examples.

Consider an individual with several job offers and many objectives relevant to the choice. Suppose one objective is to "maximize flexibility of working time" because he is the sole parent of a fourth grader. One can think of many natural attributes that may measure flexibility including the number of specific hours per week that one must be at work (i.e. they cannot be shifted to another time), the range of starting time in hours, the warning time one must give to change work time, and the number of hours per week one can choose not to work. Each of these natural attributes has built-in value judgments, namely since each of the suggested attributes are measured in hours, each attribute assumes any hour is equivalent to any other (e.g. not being able to shift out of 9:00am to noon each day is equivalent to not being able to shift out of 1:00pm to 4:00pm each day). The issue is whether these value judgments are reasonable.

A second issue is whether any of these attributes cover the range of possible consequences. It may be that the aspects of flexibility addressed by each suggested attribute are relevant to the individual. Then, perhaps all four attributes should be chosen. Alternatively, each of the natural attributes may be inferior to a constructed attribute that logically combines each of the notions of flexibility. Even when this is the case, the alternative natural attributes provide a basis to begin to construct such an attribute for flexibility.

Consider the case of an employer who is evaluating several job candidates for a position. As employers often desire a highly educated workforce, one objective is to "have a high education level". For such an objective, the attribute “average years of school completed" is frequently used. Certainly “years of school completed” is an easily understood attribute. One obvious problem, however, is that it may not cover the range of consequences, because it includes only standard school classes and does not take account of special classes, on-the-job training, workshops, many adult education offerings, and the like. It also assumes that completed years of school are equivalent across regions within a country or across countries. In addition, use of “years of school completed” includes the value judgment that formal education is all that really matters, whereas the quality of a person’s education or their willingness to learn as an adult typically reflects other considerations.

A final example concerns policy for salmon fishing on rivers in British Columbia (Hilborn and Walter, 1977). One objective was to ensure a productive fishery. Natural attributes that might measure this are the number of salmon that return to a stream each year, the average weight or condition of the fish, and the number of types of salmon and other fish that have sustainable populations in the river. The number of returning fish might be the most easily understandable attribute and the easiest to measure (through a simple count). A more complex, but perhaps more appropriate constructed attribute might measure the age distribution of salmon in the river. For example, a productive stream may be understood to have 50% young fish, 25% juveniles, 15% mature adults, and 10% in old age. A constructed attribute measuring deviations from this desired definition of a productive stream might be preferred by some scientists, but it also might be less understandable by decision makers and contain even more controversial value judgments.

Four points are made in this section. First, alternative natural attributes should usually be identified to measure a specific objective, as this process helps identify better attributes. Second, natural attributes should usually be selected when available, except when they are inferior to a single constructed attribute which comprehensively describes the consequences of an action. Third, the set of desired properties of attributes should be used to appraise the appropriateness of each of the proposed attributes. Fourth, it is usually the case that some attributes are better on some of these properties and worse on other properties, relative to other attributes. Hence, tradeoffs among the degree to which the desired properties are met need to be considered in selecting the attribute to be used for a particular objective.

6. Developing Constructed Attributes

For many objectives in important decisions, there are no obvious natural attributes. This may be the case even after trying to decompose an objective into component parts and searching for natural attributes for each component. In such cases, it is usually best to try to construct an attribute for the corresponding objective.

There is tendency to think that constructed attributes are subjective and vague, whereas natural attributes are clearer and more objective. Regarding objectivity, we have noted throughout this paper that value judgments are inherently a part of natural attributes. They definitely involve subjectivity: indeed, the broadest value judgment about a natural attribute is that a particular attribute is reasonable for the problem.

The concern that constructed attributes are often vague is valid. But this need not be the case. In this section we indicate how to build constructed scales that meet as well as possible the five desired properties of attributes discussed in Section 3.

There is also a perception that constructed scales are novel and rarely used. Part of this is because people do not recognize constructed scales, even though we come into contact with them daily. In the first minutes after being born, infants are measured on a constructed scale called the Apgar scale. This scale combines five separate features to indicate the overall health of an infant. A grade point average is a constructed scale that we are all well aware of. The wind chill index is a scale constructed to indicate the equivalent temperature with no wind as felt by your body. The recent five-color scale of the United State's terrorism alert status is another constructed scale. The Michelin rating system for restaurants that awards up to three stars for the quality of food and up to five crossed knives and forks for the level of luxury are constructed scales of restaurant quality. We are all familiar with these and many other constructed scales, but how do we construct them well? The following examples introduce five principal types of constructed scales and discuss requirements and applications of each.

Defined Levels. This basic type of constructed scale is like that shown in Table 1. Two or more levels are carefully defined to indicate possible consequences. The description of each consequence level of the attribute should be clear. Collectively, the set of consequence levels should cover the range of possible consequences related to the corresponding objective.

This defined-level constructed attribute is similar in spirit to a Likert seven-point scale. The important difference is that the levels on the Likert scale are typically either not defined or defined in an ambiguous way. Raters, who may or may not have relevant information about consequences, are often asked something like the following. "On a scale of 1 to 7, where 1 is best and 7 is worst, how would you describe the consequences of alternative A with respect to objective X?" In other situations, vague terms such as "minimum impact" are used at one end and "significant impact" at the other, with no definition of levels in between. In both cases, the consequences associated with different ratings are ambiguous and the results fail to communicate clearly.

A defined-level constructed attribute eliminates much of this ambiguity. The possible consequences of different levels should be chosen so that they cover the range of consequences and the difference in adjacent consequences is significant. To describe the consequences of an alternative using such an attribute, the judgments of people familiar with that attribute are necessary. For the power plant siting decision for which the attribute in Table 1 was developed, a community relations expert made the judgments. For one site, he could assign a consequence level of 1 meaning support. If he was uncertain concerning another potential site, he could describe possible consequences as having a 60% chance of being a level -1 (controversial) and a 40% chance of being a level -2 (action-oriented opposition).

Another example is the constructed scale of nine levels used to describe the biological impacts of building a proposed power plant at various candidate sites (Keeney and Robilliard, 1977). Prior to these assessments, a team of environmental scientists had produced environmental impact statements for each of the sites. This information was used in working with these scientists to develop appropriate definitions of the attribute levels. Also, the environmental impact statements provided significant documentation to support the judgments used to assign probabilities to different levels of biological impacts for each of the alternative sites.

Quality-Quantity Scales. There are many situations where both the quality and the quantity of an impact are relevant to adequately describe a consequence. Suppose one is concerned about the quality of water near the shoreline of different lakes in a region. Different management alternatives may affect that quality. However quality is measured, a given improvement is more significant to a lake that has more shoreline. In this case, quality might be measured by water quality indices. The quantity could be measured by the total amount of shoreline. The simplest quality-quantity measure of consequences would be a product of the total amount of lake affected times the improvement (or degradation) in water quality.

The power plant siting study described in Keeney and Robilliard (1977) had a different environmental objective that concerned both quality and quantity. It was to minimize the impact on salmon in streams where cooling water would be withdrawn at alternative power plant sites. Some of those streams had an annual salmon run in the thousands, whereas the largest source of cooling water was the Columbia River, which has a salmon run of a million salmon annually. The loss of 5,000 salmon in a river that has a total run of 10,000 is much more significant than the loss of 5,000 salmon in the Columbia River. On the other had, a 20% loss of salmon in the Columbia, which is 200,000 salmon, is much more significant than a 20% loss in a river with 10,000 salmon, which is a loss of 2,000 salmon.

There are two ways that one might develop useful constructed attributes for this situation. One is to use a value model as described below. The other way is to weight quality by quantity. A straightforward way to do this is to first assess the relative value of the loss of a single fish in rivers with different annual salmon escapement. These numbers would then be multiplied by the number of fish lost in the rivers of that size to get an equivalent number of fish lost. As a standard, one may assign 1.0 as the value of a salmon lost in the Columbia River. Then, if fish lost in a river with 10,000 escapement were deemed 15 times as important for the health of the fishery, 15 would be the weight assigned of such a river. If that river was expected to lose 200 fish, the total equivalent fish lost would be 15 times 200, which equals 3,000 Columbia River salmon. To incorporate nonlinearities and/or thresholds, a more complex evaluation function (e.g. a multiattribute utility function) addressing the same attributes should be used (Keeney,1992).

Value Models. An example of using a value model to develop a constructed attribute comes from the analysis of alternatives to improve the reliability of the British Columbia Hydro electricity system. One of the objectives concerned minimizing the likelihood of blackouts. The significance of a blackout depends on its duration and on how many people are affected. The former might be thought of as referring to the quantity of the blackout and the latter to its quality. Based on the judgments of planners at BC Hydro, Keeney et al. (1995) constructed an attribute using a simple value model that was the product of the outage duration t and the number of residences impacted n as an attribute for blackouts. Thus, if b measures the blackout, it is defined as $b = tn$. Note that this simple value model is just the continuous version of weighting quality by quantity and vice-versa. One might interpret this weighting of quantity and quality as decomposing the overall concern into two parts and using natural scales for each of those parts. Once the two are combined by multiplication, a relatively simple constructed scale is created.

An example of a more complex value model for developing a constructed attribute comes from the measurement of trauma severity as an element for evaluating hospital emergency room quality. Consider a situation where the survival rate of individuals entering an emergency room in one hospital is 80% and in another hospital 95%. A simple interpretation would suggest that the latter hospital was much better. However, if the trauma was more severe of those entering the former hospital, then it could be providing the better service. In any case, adjusting for the severity of trauma of individuals entering an emergency room is necessary to appropriately examine the relative quality of services. Working with a trauma surgeon who had experience in several emergency rooms, a value model was developed that incorporated seven major trauma concerns: ventilation, circulation, central nervous system, internal organs, renal function, muscular and skeletal system, and burns (Fryback and Keeney, 1983). Natural attributes were developed for measures of most of these types of trauma and then integrated together using the value judgments of the trauma surgeon. The model, which included many nonlinear relationships, was subsequently used to make comparisons across sites and to suggest improvements at different emergency rooms.

Weighted Scales. In many situations, a reasonable constructed attribute is a simple weighted scale of different impacts. The scale to determine a grade point average is one example. Another familiar example is the full-time equivalents for number of employees. There are many government decisions that have the creation of jobs as one objective. Some of these jobs might be half time, some might be three-quarters time, and some might be full time. If we weight the number of full time jobs by 1.0, the number of three-quarter time jobs by 0.75 and the number of half time jobs by 0.5, then summing up the number of jobs created in each category multiplied by the weights gives the number of full time jobs created.

Another example comes from a study to examine integrated resource plans for British Columbia Gas (Keeney & McDaniels, 1999). One of the objectives was to minimize social disruption. The types of disruptions of citizens were categorized from noise to odor to traffic congestion to fear and worry (from having a natural gas facility nearby). The relative seriousness of a day of these different types of impacts was assessed by a team of stakeholders and used to calculate total equivalent days of social disruption of different alternatives.

The general case can be illustrated using the objective of minimizing injuries from car crashes. There will be different types of injuries categorized by seriousness, and it is appropriate to have the level of seriousness in each category be distinct from that in another category. Assume that five levels of seriousness have been categorized, namely levels A, B,…,E. Then one needs to make value judgments about the relative seriousness of each of those levels. Suppose level A is the least serious and is arbitrarily assigned a weight $w_A = 1.0$ to indicate its seriousness. If the seriousness of an injury of B is three times as important, then the weight for injury B is $w_B = 3.0$. Suppose the other weights are found in a corresponding manner. Then the overall impacts on injury, defined as x, can be calculated using

$$x = w_A \text{ (number of type A injuries)} + w_B \text{ (number of type B injuries)} + \ldots + w_E \text{ (number of type E injuries)}, \quad (1)$$

where x is the equivalent number of type A injuries for an alternative.

The constructed attribute in (1) can be adjusted for interpretation as the equivalent number of type B injuries or any other type of injuries. To do this, we simply divide x by the weight of the corresponding type of injury. For instance, x/w_E would give us the equivalent number of type E injuries from an alternative.

The choice of how to represent these injury consequences should depend on which of the component injuries is generally the biggest contributor to x. If 70% of the equivalent impacts were due to injuries of type C, then it would probably be best to use the constructed attribute of equivalent injuries of type C to describe consequences. This case would require less of an adjustment due to construction to obtain an equivalent number of injuries.

<u>Pictures</u>. For some decisions, such as those that involve siting transmission lines, one objective might be to minimize visual degradation. If transmission lines are blocking a beautiful view along a highway, then there is visual degradation. The more people that use the highway, the greater the degradation. This example has a component of the quality-quantity problem. In this case, the quantity can be indicated by the natural attribute concerning the number of people using the highway. The quality of the impact, however, might best be illustrated with a set of pictures created to indicate what the scenery might be from different locations if different alternatives are pursued. For instance, a transmission line much further from the road would presumably have less visual impact.

Standards for using a visual attribute are essentially the same as for a defined level scale. One develops a set of pictures that cover the range of impacts from best to worst. There should be enough pictures such that each significantly different level is characterized by a different picture.

A very ingenuous use of pictures for an attribute concerned the evaluation of treatments for children with the common congenital problem of cleft lip and palate (Krischer, 1976). One objective was to minimize facial physical disfiguration. For this attribute, Krischer began with facial pictures of several children approximately 10 years old. The faces showed different amounts of disfiguration subsequent to treatment for cleft lip and palate. However, since the hair, eyes, cheeks, and chin are not affected by the treatments, they were not used on the faces for the constructed attribute. A professional artist drew those features, which were then common to all faces. The part of the photos showing only the nose and mouth areas were superimposed on this sketch. The superimposed part was a triangle from the top of the nose to both sides of the lower jaw connected with a line cutting across the chin just below the mouth. As noted in the article, this creative application of a constructed attribute helped to influence clinical decision making by investigating potential differences among the preferences of clinicians and families for different treatments for cleft lip and palate.

7. Conclusion

In conducting an analysis of any important decision, a number of objectives are typically relevant. Attributes are necessary to indicate the degrees to which each of these objectives are met by the various alternatives. The thoughtful choice of attributes clarifies the meaning of each objective, provides for a useful description of the consequences of each alternative, and facilitates an insightful evaluation of alternatives.

This paper outlines a systematic process for appraising possible attributes and for selecting the best attribute in any given situation. In general, if a natural attribute can be found, it should be used. If no single natural attribute is appropriate, then one should attempt to find a set of natural attributes that adequately describe the consequences pertaining to a specific objective. When that can't be done, constructed attributes that directly measure the impact should be developed. This usually is possible, albeit with effort and careful thought. Too often, people give up at this stage or don't even consider developing a constructed attribute. Instead, they end up neglecting a relevant objective, saying it can't be measured, or choosing a proxy attribute that only indirectly indicates the degree to which the corresponding objective is met. Such a choice de facto discards much of the potential insight that might be gained from a thoughtful analysis.

There are two main reasons for most of the shortcomings of attributes. The first is selecting a natural attribute without sufficient care. Often the value judgments built into this natural attribute are inappropriate or the attribute really is a proxy and therefore it is not very relevant. The second reason, which most often arises with constructed attributes, is that inadequate attention is given to do a thorough job. Recognizing that a choice exists among different types of constructed attributes and paying close attention to the five desired properties of attributes – unambiguous, comprehensive, direct, operational, and understandable – will help one to build constructed scales with the appropriate thoughtfulness and attention.

The overall message of this paper is simple. Selecting attributes to measure the achievement of objectives is an important task worthy of time and effort. This basic message is true regardless of who is involved in the selection of attributes (e.g., an individual decision maker, a technical advisory committee, or a representative body of diverse stakeholders) or the particulars of the specific problem under consideration. Thinking carefully about each attribute is an essential part of the analysis. As in any decision (in this case, choosing attributes), alternatives (in this case, alternative attributes) should be considered in light of the relevant tradeoffs (in this case, among the five desirable properties) unless the choice of one attribute is completely obvious.

Good attributes are necessary to describe how well each of the alternatives under consideration satisfies the objectives of concern and to make reasoned value tradeoffs between those objectives. Without good attributes, useful insights from analysis are limited. The theory and procedures and practice are such that good attributes can almost always be found if one considers this a problem worthy of time and effort.

Acknowledgement
All the referees' suggestions were helpful. One referee had numerous great suggestions, and he or she will clearly recognize the impact of these suggestions. Thank you. Also, support from the National Science Foundation (Grant SES-0114924 to Decision Research) is gratefully acknowledged. The views expressed in the paper are those of the authors and do not necessarily represent the views of the National Science Foundation.

8. References

Camerer, C. & Weber, M. 1992. Recent developments in modeling preferences: uncertainty and ambiguity. Journal of Risk and Uncertainty, 5: 325-370.

Ellsberg, D. 1961. Risk, ambiguity, and the Savage axioms. Quarterly Journal of Economics, 75: 643-699.

Fiske, A. & Tetlock, P. 1996. Taboo trade-offs: reactions to transactions that transgress spheres of justice. Political Psychology, 18: 255-97.

Fryback, D. & Keeney, R.L. 1983. Constructing a complex judgmental model: an index of trauma severity. Management Science 29: 869-883.

Gold, M., Siegel, J., Russell, L. & Weinstein, M. 2000. Cost-Effectiveness in Health and Medicine. Oxford University Press, New York.

Gregory, R. & Failing, L. 2002. Using decision analysis to encourage sound deliberation: Water use planning in British Columbia, Canada. Journal of Policy Analysis and Management, 21 (3): 492-499.

Gregory, R. 2003. Using decision analysis to allocate funds to inner city schools. Manuscript in preparation.

Hilborn, R. & Walters, C.J. 1977. Differing goals of salmon management on the Skeena River. Journal of the Fisheries Research Board of Canada, 34: 64-72.

Hsee, C. 1996. The evaluability hypothesis: An explanation of preference reversals between joint and separate evaluations of alternatives. Organizational Behavior and Human Decision Processes, 46: 247-257.

Keeney, R.L. 1982. Decision analysis: An overview. Operations Research, 30: 803-838.

Keeney, R.L. 1992. Value Focused Thinking: A Path to Creative Decision Making. Harvard University Press, Cambridge, Mass.

Keeney, R.L. & McDaniels, T. 1999. Identifying and structuring values to guide integrated resource planning at BC Gas. Operations Research, 47: 651-662.

Keeney, R.L. & Sicherman, A. 1983. Illustrative comparison of one utility's coal and nuclear choices. Operations Research, 31: 50-83.

Keeney, R.L. & Robillard, G. 1977. Assessing and evaluating environmental impacts at proposed nuclear power plant sites. Journal of Environmental Economics and Management, 4: 153-166.

Krischer, J. 1976. Utility structure of a medical decision-making problem. Operations Research, 24: 951-972.

McDaniels, T.L., Gregory, R. & Fields, D. 1999. Democratizing risk management: Successful public involvement in local water management decisions. Risk Analysis, 19: 497-510.

Merkhofer, M.W. & Keeney, R.L. 1987. A multiattribute utility analysis of alternative sites for the disposal of nuclear waste. Risk Analysis, 7: 173-194.

Seelye, G.Q, & Tierney, J. 2003. EPA Drops Age-Based Cost Studies. New York Times, Late Edition, Section A, May 3, 34.

Slovic, P. 1987. Perception of risk. Science, 236: 280-85.

U.S. Department of Energy. 2000. Paths to closure: Status Report 2000. Washington D.C., Office of Environmental Management (DOE-EM-0526).

Table 1.
Constructed Attribute for Public Attitudes[a]

Attribute Level	Description of Attribute Level
1	*Support:* No groups are opposed to the facility and at least one group has organized support for the facility.
0	Neutrality: All groups are indifferent or uninterested.
-1	*Controversy:* One or more groups have organized opposition, although no groups have action-oriented opposition. Other groups may either be neutral or support the facility.
-2	*Action-Oriented Opposition:* Exactly one group has action-oriented opposition. The other groups have organized support, indifference, or organized opposition.
-3	*Strong Action-Oriented Opposition:* Two or more groups have action-oriented opposition.

[a] This attribute was used in Keeney and Sicherman (1983) in an analysis to evaluate potential power plant sites.

Figure 1. A Flow Chart for Selecting Attributes